

Semi-Annual Progress Report

July 1994

Mark R. Abbott

College of Oceanic and Atmospheric Sciences

Oregon State University

MODIS Team Member, Contract # NAS5-31360

Task Objectives

The objectives of the last six months were:

Continue to review plans for EOSDIS and assist ECS contractor

Continue development of local information management system

Continue analysis of sun-stimulated fluorescence data collected off northern California

Prepare an Algorithm Theoretical Basis Document for planned at-launch data products and participate in ATBD review.

Work Accomplished

- Project Data and Information System Plans
 - ECS Scenario Development

Documents and standards continue to be produced by the ESDIS Project and by the EOSDIS Core System (ECS) contractor. My team has participated in these reviews, especially in preparation for the recently-completed System Design Review (SDR). Our activities were divided into three components: development of a user scenario, review of the SDR documents, and participation in the Data Processing Focus Team (DPFT).

Working with personnel from HITC, I developed a scenario to blend chlorophyll fields derived from MODIS and MERIS. Beginning in 1998, there will be more than one ocean color sensor in orbit. As each sensor

will have different operating and sampling characteristic, it will not be a simple matter to merge all of the available observations of chlorophyll into a single, consistent product. The scenario involved retrieval of sensor operating and calibration characteristics in order to determine the associated errors with each data source. An optimal assimilation scheme would be developed to combine the data into a consistent estimate of global ocean chlorophyll with estimated errors. This scenario will be refined over the next few months and will be used to help HITC size the ECS.

- ECS System Design Review

We devoted a significant amount of time to the review of the SDR documents produced by the ESDIS Project and HITC. In general, the SDR went well. I provided the following Review Item Discrepancies:

- From the presentations, it appears that a key element is the need to develop some standards in the area of metadata, attributes, vocabulary, etc. in collaboration with the science community. Without such standards, development will proceed slowly. Present plans to come to closure on these and related issues are vague, and depend on strong Project involvement. Studies and working groups have been formed for the past 20 years, with little progress. Without such agreement, ECS is at great risk in terms of schedule.
- Presently plans rely heavily on NSI and the Internet for data access and delivery. As the federal subsidy for the Internet Access Providers disappears, there is the possibility that science users may need to bear the direct costs for Internet access, perhaps on a per packet basis. If network costs significantly increase and particularly if these costs are passed to the user based on network usage, then the relative balance between electronic delivery and fixed media may change. This will affect ECS costing. Also, users may require different methods for browsing or even access (such as high speed dial-up) to avoid perceived Internet costs. HITC should explore sensitivity of system cost and architecture to changes in network pricing policies.
- It is not clear how passwords will be managed in this system. There is the potential (based on the user model) for hundreds of thousands of user accounts, each of which must be

unique. When questioned, there appeared to be no solution to this serious management issue. What real-world systems are in place that have confronted this challenge? HITC should document a solution to this problem and its potential technical requirements that are consistent with the ECS performance spec in terms of response time, etc.

- What is the approach to updating distributed and/or dependent objects? Will such upgrades or enhancements happen autonomously? How will the system maintain inheritance control? The complexity of the system will make this a difficult problem. Answers to oral questions were not adequate; Contractor recognizes problem but has no solution.

- The present "topology" is considering a variety of LANs from switched ethernet, to FDDI, to ATM, and to FCS. These approaches differ in their fundamentals (switched vs. routed, etc.) and will significantly affect system design (both hardware and software). It is surprising that the choices have not settled down yet at this stage in the design. HITC should settle on a design, or at least one that minimizes risk if "solution" needs to be changed later.

The results of the SDR were far superior to the Systems Requirements Review last fall. We feel that HITC has a reasonable architecture, although there clearly are some shortcomings at present. However, the most challenging issue is cost. Can HITC meet the ESDIS requirements in a cost-constrained environment? This question has not yet been answered.

- Local Scientific Compute Facilities
 - Advanced Networking

The joint project with Otis Brown at the University of Miami finally began in late June. Under the auspices of the Naval Research Laboratory, MCI installed an ATM link between our sites and NRL in Washington, DC. The plan is to supply an optimal interpolation model running on the CM-5 at OSU with global AVHRR data collected at Miami. These data will be displayed back to Miami. At present, we are still testing the basic network connectivity. Some kernel modifications are needed for the CM-5 to accommodate an ATM link, and FORE Systems is continuing to work on this problem.

The NIIT project which links OSU with Berrien Moore at University of New Hampshire has been on hold for the past 6 months. A proposal was made to HITC to support development of SCF to SCF links to test the software and hardware requirements for interdisciplinary research. We are in the final stages of proposal development, and we expect to begin this work this fall.

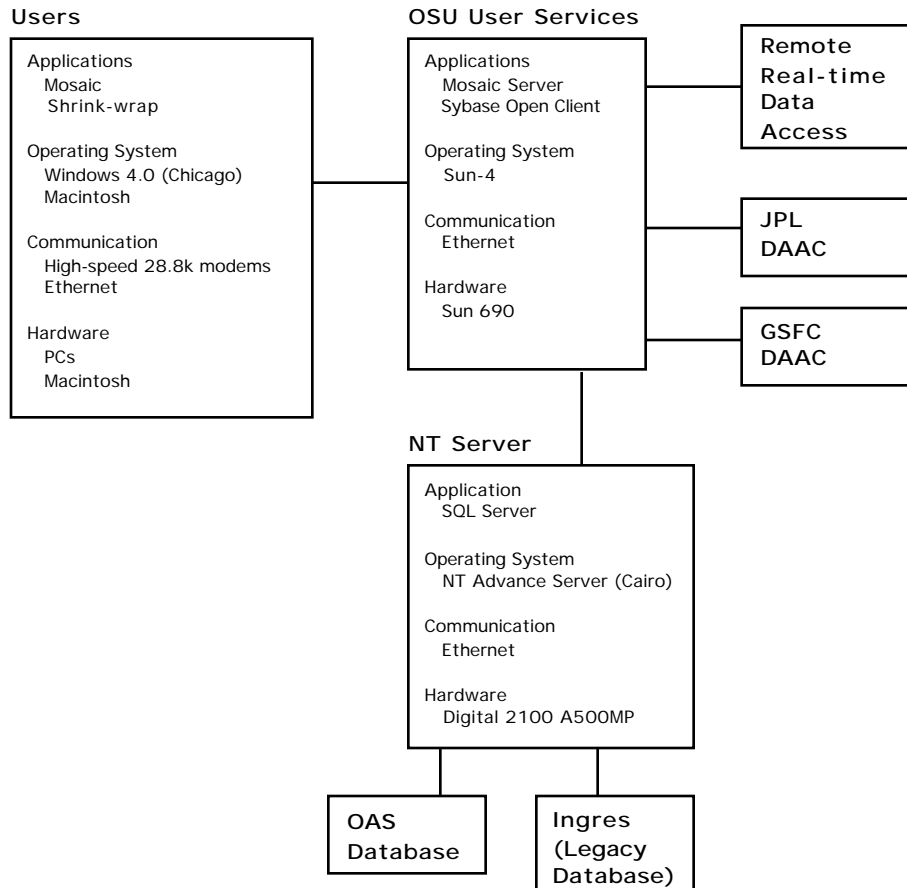
- Information Systems Development

In the past data centers were used to provide access to large data sets. However, the transition from data centers that contain most of the compute power and data storage to distributed centers linked by a network is nearly complete. The next stage will be the integration of small systems that will have storage in the range of 10 gigabytes and compute power in the range of 50 Mflops. Such capabilities were the exclusive province of data centers only a few years ago. With hundreds of such "centers" accessible through the network, the challenge is to provide simple, integrated access.

We are developing a context-based query system that satisfies the following high-level architectural requirements:

- Enable use of off-the-shelf applications as much as possible.
- Reduce maintainability cost.
- Provide flexibility for changes in data structure.
- Provide context-based searches.
- Support flexible means to use data.
- Use existing hardware as much as possible.
- Develop a low cost solution.

To satisfy these requirements, we plan to use the following client/server architecture.



We will import our flat-file data into a relational database that will reside on an NT Server. Users will then run shrink-wrap applications on PC and Macintosh desktop clients. Due to the inherent qualities of the NT Server, these shrink-wrap applications will be able to directly access data stored in the relational database without the need for any programming. This will allow users to use familiar applications (e.g., Microsoft's Excel and Research Systems' IDL) to access and manipulate data. In addition, since Mosaic supports SQL queries, users will be able to access our database via Mosaic running on a PC, Macintosh or UNIX workstation.

The benefit of this approach is threefold. First, users will no longer be limited to predefined data sets. Instead users will be able to make their own queries and obtain just the information they need. Second, users will be able to use popular shrink-wrap applications to directly manipulate the data without the need for a programmer. For instance, users could directly access the database from within Excel and then create charts based on this data. Third, users will no longer need expensive UNIX workstations to access oceanic data. Instead, low-cost PCs and Macintoshes can be

connected via high-speed modems to the NT Server. In addition, users can take advantage of the low-cost shrink-wrap software that is sold for these desktop systems.

This client/server architecture is based on the following three major architectural requirements: (1) the system will utilize the embedded communication features of Microsoft's NT Advance Server and Windows 4.0, (2) our data will be stored in SQL Server (a relational database) running on an NT Server, and (3) the interface between the client applications and the server database will be via Microsoft's Open Database Connectivity (ODBC).

i. NT Advance Server and Windows 4.0

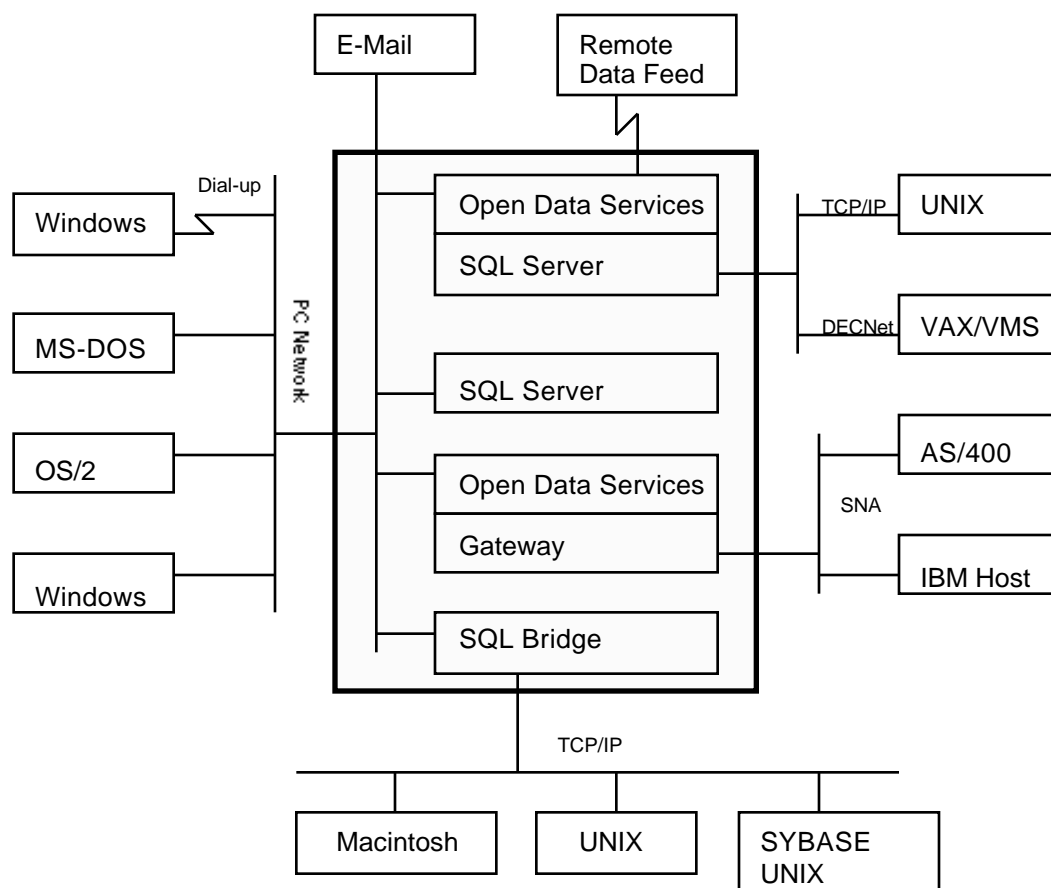
When NT was first conceived, Microsoft wanted to provide an operating system with UNIX-like features (networking, multiprocessing, etc.) that would provide a familiar Windows environment to users that were uncomfortable with the complexities of UNIX. Microsoft hoped NT would find wide acceptance among the broad majority of computer users that lie outside the technical and university community. As part of Microsoft's ongoing upgrades of its existing Windows NT, Windows 3.1 and DOS 6.2 products, Microsoft will soon introduce their next generation NT Advance Server (Cairo) and Windows 4.0 (Chicago). Though these operating systems maintain the Windows look and feel, they use a new application development framework that offers the following major benefits:

- The number of people familiar with the Windows environment is larger than those familiar with UNIX. Windows 4.0 has the potential to replace the 50 million plus versions of Microsoft Windows.
- NT inherits a large legacy of Windows applications that dwarfs by far the number of applications that run on its nearest competitor (Sun's Solaris OS).
- Windows 4.0 will have complete integrated TCP/IP support. This will allow users to run Mosaic (or similar such applications) directly on their desktop using a high-speed modem dialed into an NT server. This will enable users at home or in the schools to make context-based queries and promptly receive the requested information over the network.
- Data can be stored on a low-cost high-performance scalable multi-processing NT server.

We propose to develop our own system using one of several vendors' multiprocessor NT servers (Digital Equipment, Hewlett-Packard, or Compaq). We will use this four processor server to run our relational database (Microsoft's SQL Server). Our server will support PC (running Windows 4.0) and Macintosh clients. With the installation of Sybase Open Client for Macintosh, any Macintosh can query our NT server. Since most K-12 users have either PCs or Macintoshes, this approach will allow these users access to our data. UNIX users will be provided access via Mosaic.

ii. SQL Server Relational Database

Microsoft SQL Server is a high-performance network database management system designed to help companies integrate, manage, and access data anywhere in an enterprise. SQL Server can act as a central hub through which enterprise data is transferred to the desktop. Users can therefore work with corporate data using the same graphical applications they use to work with personal data.



Connections to the many environments shown in the preceding figure can be through facilities provided with SQL Server, or through a wide variety

of third-party gateway solutions. Since SQL server provides built-in connectivity to existing Ingres, Oracle and Sybase databases (e.g., JPL DAAC and GSFC DAAC), our system will provide public access to NASA's Earth and space science remote sensing databases.

One of the greatest benefits of the SQL Server design is that it supports distributed data management, making truly distributed applications possible. The SQL Server distributed update capability allows databases on multiple SQL Server installations to be updated with a single transaction. This ability to scale a system in response to data server requirements provides substantial flexibility, permitting access to geographically or functionally dispersed data.

In addition, SQL Server offers the following features:

- SQL Server-compatible gateways enable applications to access and update data from other systems (e.g., DB2, CICS, Oracle, SQL/400, Ingress).
- Microsoft's SQL Bridge allows complete interoperability between Microsoft and SYBASE SQL Server environments. PC-based applications can access SYBASE SQL Server on UNIX and VMS. And client applications written for UNIX, VMS, and Macintosh clients can use Microsoft SQL Server on PC networks.
- SQL Server supports remote stored procedures providing a transparent, distributed server-to-server communication mechanism between SQL Server and Open Data Services applications. A special type of stored procedure, known as a trigger, can be associated with particular tables, and automatically initiated whenever attempts are made to modify these tables. Triggers can protect data and take actions based on changes to a table. For instance, triggers can perform quality assurance tests on incoming data and automatically notify users via email of any data degradation.

iii. Open Database Connectivity (ODBC)

This relational database (SQL Server) running under NT will allow the majority of desktop computers to make context-based queries of oceanic and atmospheric data. These queries can be made in two ways. First, users can use off-the-shelf applications to directly query our data without the need for any programming. Second, users can connect to the Mosaic server we intend to develop. Since Mosaic supports SQL queries, any authorized user on the Internet will be able to easily query our database.

In either case, the key to providing this connectivity is our plan to standardize on Microsoft's Open Database Connectivity (ODBC).

ODBC enables client applications in the Windows environment to connect to many different databases directly as well as through gateways, which are specialized applications that translate database queries from one format to another. Based on the call-level-interface specification proposed by the SQL Access Group (a standards group made up of various database application vendors), ODBC was designed to work with a wide range of DBMS systems from different vendors. Instead of using a different API to access each database, or developing a single API that taps only the lowest common denominator of a number of different databases, ODBC's API enables applications to take advantage of a database's full capabilities. In essence ODBC can be thought of as an abstraction of SQL. Subsequently any application that can talk ODBC can directly access data stored in a relational database. This allows users to access data stored in a relational database from within their off-the-shelf application without the need for a programmer. Thus, by way of ODBC, we can provide integrated access to data through the desktop applications required by the user.

- Future Directions

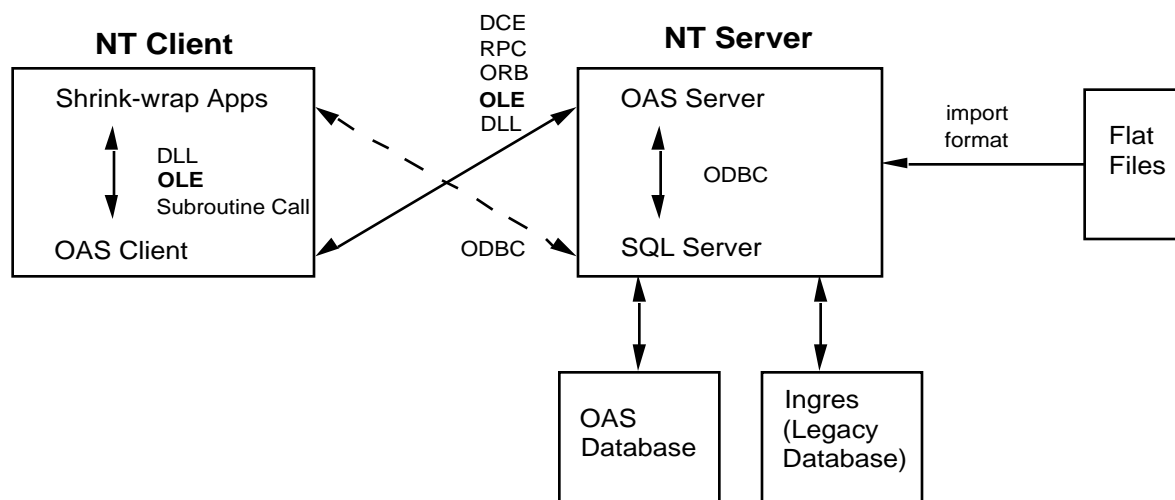
Though this relational database paradigm offers connectivity benefits, the relational model has its shortcomings. The original objectives of the relational model were simplicity, data independence, and rigor. But the relational model of viewing data as if they were formatted into tables is so simple that it cannot explicitly capture the semantics needed for non business applications. For instance, there is no way to explicitly express the notion that one table is composed of other tables. There is only a very weak notion of linkages between tables, thus the necessity of much join processing. Furthermore, the data types in the relational model are constrained. Typically, the relational data types are integer, decimal, float, character, and string. Sometimes, types like money, date, time and so forth are supported. But there is no facility for users to define their own data types.

The model supported by an object-oriented database (ODBMS) is more powerful in its capability of explicitly expressing application semantics. For instance, the object-oriented paradigm supports the notion of inheritance. Thus, you can explicitly express the notion that one set of data is composed of other sets of data. Furthermore, with an ODBMS the program deals with persistent and transient objects as one uniform space. No longer is the database considered separate from the program. This

greatly simplifies programming. Furthermore, the object database design is a fundamental part of the overall application design process. The object classes used by the programming language are the classes used by the ODBMS. Because their models are consistent, there is no need to transform the program's object model to something unique for the database manager. Subsequently, there is no need to try to figure out how to represent real-world objects within the confines of tables in such a way that results in good performance and preserves data integrity.

Due to these limitations of the relational model, we plan in the future to develop an object-oriented client and server that will reside on top of the relational database. This architecture is illustrated in the following figure.

Oceanic & Atmospheric Sciences (OAS) System Architecture



- Data Analysis and Interpretation
 - Ocean Color White Paper

A strategy for developing a 20-year time series of ocean color observations was published as SeaWiFS Technical Memorandum 17. It was distributed to the SeaWiFS and MODIS science teams as well as to members of the EOS Payload Panel.

- Journal Publications

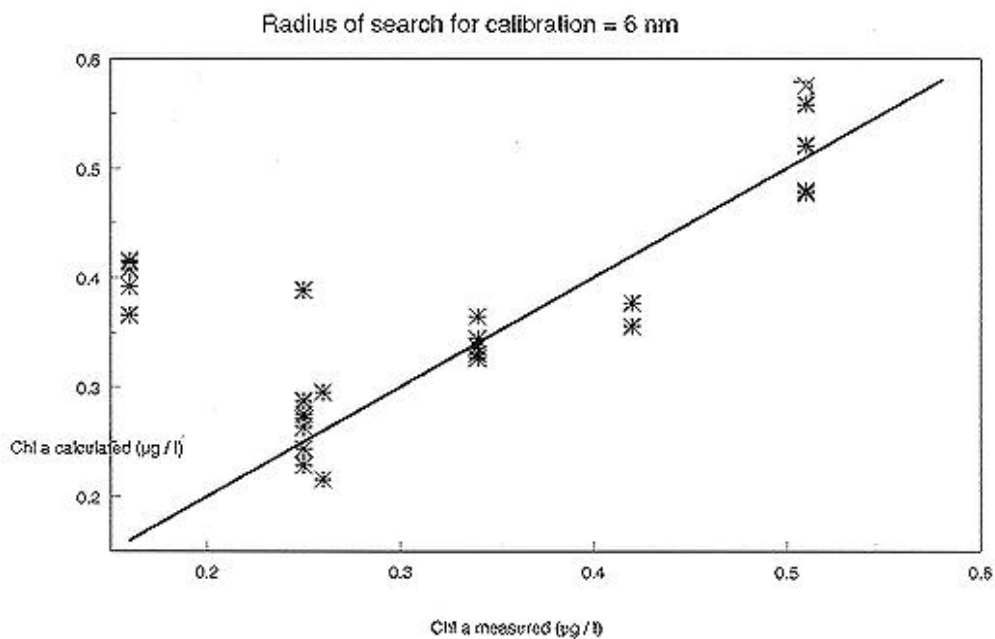
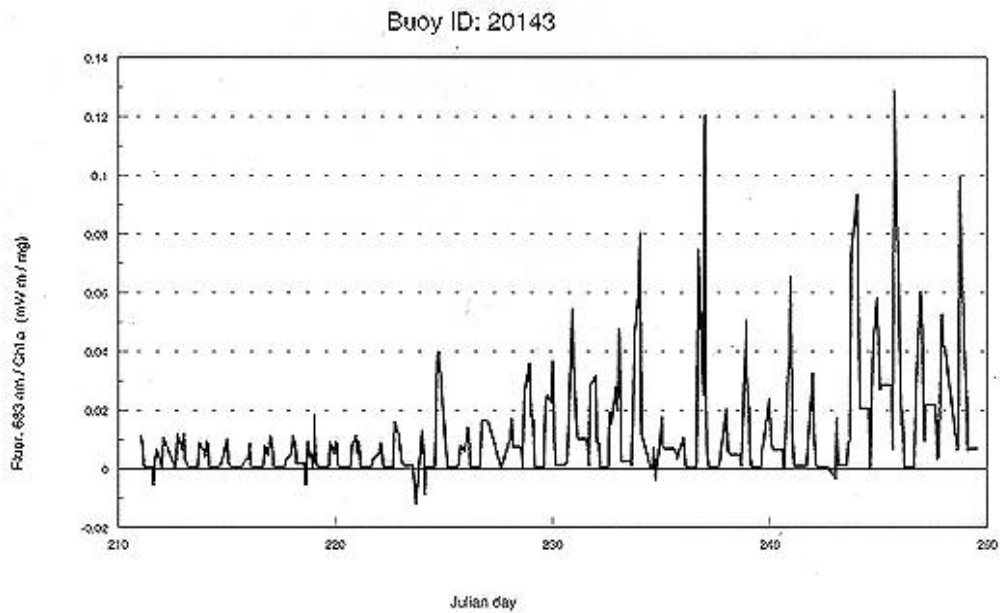
A paper describing the scales of bio-optical variability off northern California was accepted for publication in the Journal of Geophysical Research. It will appear later this year in a special issue on bio-optics.

- Analysis of Ocean Drifter Data

I have deployed 20 near-surface bio-optical drifters in the California Current as part of ONR's Eastern Boundary Current study. Five drifters were deployed this year; four more should be deployed in the next

month. Dr. Ricardo Letelier began his postdoctoral research with me in mid-May, and he has been focusing on analysis of these data.

The top figure is fluorescence per unit chlorophyll, where chlorophyll has been estimated using the radiance ratio method of Clark. The fluorescence data have been corrected for solar backscatter. Aside from the diel signal, we note a pronounced shift in the fluorescence:chlorophyll ratio as we move offshore. This is not unexpected; low productivity water offshore should have a higher ratio as more light energy is lost to fluorescence rather than being used in the photosynthetic process. The bottom figure shows a comparison of the radiance ratio estimates of chlorophyll with ship-based measurements of chlorophyll. The relationship is surprisingly good, considering the strong mesoscale and small-scale variability in this region. We interpret the cluster of point at the low chlorophyll end to be the result of the drifter being on one side of the coastal transition front and the ship being on the other.



Dr. Letelier will continue to analyze these data, and the results will be used to develop field and lab experiments. Our ultimate objective is to understand the variability in the fluorescence:chlorophyll signal and how this variability can be used to quantify the physiological state of the phytoplankton. This information will be used to improve estimates of primary productivity in the MODIS era.

- Algorithm Theoretical Basis Document

I have completed my Algorithm Theoretical Basis Document (ATBD) in early March. Reviews were received in mid-April, and they were all

positive about my approach. The only concern is the interpretation of the fluorescence signal and its relationship to primary productivity. This concern was also voiced at the panel review. I noted during the review that most of my MODIS research is focused on lab and field studies to improve our understanding in this area.

The panel expressed some concern on the quality control/assurance process. I will explore the implementation issues related to QA over the next several months. I have not received a panel report yet so I can only address the few comments made during the presentation.

Anticipated Future Actions

I will work with HITC on various aspects of the ECS. This will involve user scenarios as well as further definition of processing requirements. My delivery of beta code will be coordinated with the MODIS researchers at the University of Miami.

In the area of advanced networking, we will continue our joint work with the University of Miami and NRL. We have ported most of the optimal assimilation code to the CM-5. The network is still undergoing basic connectivity tests, but these should be completed soon. We will then move global AVHRR data from Miami to OSU to run through the OA models. Our other network experiment with Berrien Moore to test SCF/SCF interactions should begin this fall with HITC support.

Our information management system using SQL Server and Windows Advanced Server will be running this fall. We will begin to develop client applications to merge our analysis tools with the data management system.

Dr. Letelier will use the results of the drifter analyses to design lab and field experiments. He will participate in a Long Term Ecological Research cruise from the Antarctic Peninsula during December 1994. He will collect photosynthesis/irradiance information and compare these measurements with sun-stimulated fluorescence data. The fluorescence data will be collected from two bio-optical drifters as well as from a tethered spectroradiometer from Satlantic.

We will also deploy a Satlantic spectroradiometer on a mooring in the Antarctic Circumpolar Current using EOS/IDS funding. The instrument will also measure fluorescence. We hope that the mooring will still be in the water when SeaWiFS is in orbit.

We are trying to acquire other sun-stimulated fluorescence data sets, notably the Hawaii Ocean Time Series which is part of the US JGOFS program.

Lastly, we intend to work with researchers at OSU to purchase a “pump and probe” fluorometer to study primary productivity. This device will provide valuable information and can be used either in the lab or at sea.

Problems and Solutions

No significant problems have been encountered during this reporting period.